Introducing Machine Learning to Non-Majors Using An Interactive Machine Learning Tool

Amal Altuwaiyan **Mahalakshmi Sundaresan**

ABSTRACT

The rapid growth in the application of machine learning (ML) has led to an increasing effort to prepare non-major students to use ML in their work. With ML skills combined with their domain knowledge, non-majors can contribute to developing new innovative applications of ML across different domains. In this study, we studied how Google Teachable Machine, an interactive machine learning tool, shapes non-majors' understanding of basic ML concepts, and how Google Teachable Machine helps non-majors train, evaluate, and refine ML models. Four non-major undergraduate students participated in this study. They used Google Teachable Machine to train ML models. Interviews and observations were conducted. Our findings reveal that Google Teachable Machine facilitates the construction of viable mental models of basic ML concepts and that non-majors can train highly accurate ML models using Google Teachable Machine. In addition, Google Teachable Machine positively affected non-majors' perceptions, interest, and attitudes towards learning about ML.

INTRODUCTION

The rapid growth in the application of machine learning (ML) has led to an increasing effort to prepare Computer Science (CS) students and non-major students/practitioners, those with no to little background in CS, to use ML in their work, resulting in an ever-larger number of ML courses, in-person university courses and online courses, with enrollments in the millions [24]. Despite this rapid growth in the application and teaching of ML, research on how to effectively teach ML is still sparse, especially for non-majors. The little research that exists on ML education for non-majors spans across teaching ML to practitioners [8], non-major undergraduates [26, 27, 10], youth [32, 31, 30, 16, 6], and kids [12, 22, 13]. These studies show that non-majors with limited background in CS and statistics are able to understand and implement ML [26, 27, 10, 28], but further research is still needed to study and develop educational tools that could support ML education for non-majors [26, 27].

In non-academic efforts, several interactive machine learning (IML) tools have been developed to make ML more accessible to those with little to no background in programming and statistics (e.g. Google Teachable Machine [11], Microsoft Azure Machine Learning Studio [18], and RunwayML [1]). To our knowledge, no published research examines the efficacy of such popular IML tools in introducing ML to nan-majors. Motivated to widen the literature on ML education for non-majors, we attempt to investigate how effective is Google Teachable Machine in helping non-majors understand and effectively apply ML concepts. In this paper, we define non-majors as college students with no to limited background in CS and statistics. Teachable Machine (Figure 1) is a novice-friendly IML tool that was developed to make ML more accessible to those with limited background in CS and statistics. With Teachable Machine, users can create ML models using images, audio, and poses without any programming required.

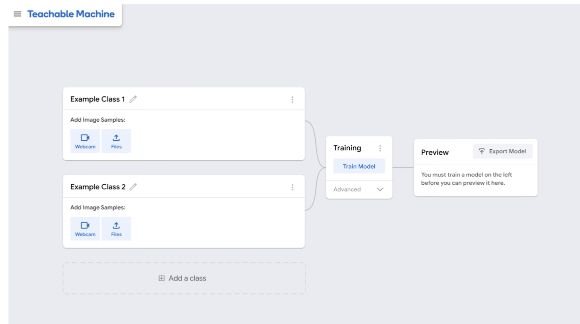

Figure 1. Google Teachable Machine User Interface

To our knowledge, no previous research has investigated user experience while learning with a popular IML tool or has attempted to guide effective use of such tools to teach ML to non-majors. We believe that such work will help educators select the appropriate tools to support ML teaching and learning. Such work will also inform the design and development of IML tools to support better understanding and application of ML for a wider audience. We selected Teachable Machine in this study for two reasons. The first reason is Teachable Machine's increasing popularity among different populations, including creators and educators who are interested in introducing ML to their students. This popularity can eventually lead to a larger community that shares projects, lessons, and discussions. The presence of such a community is vital in supporting learning with any educational tool. The second reason is Teachable Machine's beginner-friendly user interface that can be adapted to teaching ML for kids, youth, or college students. Thus, Teachable Machine could be used to scaffold ML teaching at different educational levels.

ACM ISBN 978-1-4503-2138-9.
DOI: 10.1145/1235

In this exploratory research, we attempt to answer the following questions:

RQ1:How does Google Teachable Machine, an interactive machine learning tool, shape non-majors' understanding of basic ML concepts?

RQ2: How does Google Teachable Machine help non-majors train, evaluate, and refine ML models?

We hope that the investigation of the efficacy of using Teachable Machine to introduce ML to non-majors will contribute to informing both the design and use of IML tools for non-majors. The breadth of applications of ML across different domains imposes the need to introduce ML to non-majors. With ML skills combined with their domain knowledge, non-majors can contribute to developing new innovative applications of ML across different domains [24].

RELATED WORK

Machine Learning Education For Non-Majors

As machine learning is a relatively new field, research on the teaching and learning of ML is scarce, especially for non-majors. Amy Ko [14] writes," We still know little about what students need to know [about ML], how to teach it, and what knowledge teachers need to have to teach it successfully ... If we knew these things, and we had ways to train teachers about these things, I hypothesize that students would learn how to apply machine learning more efficiently and effectively". Shapiro and Fiebrink [24] carry these concerns into the study of ML education for non-majors, they ask, "What do tomorrow's practitioners in other domains—across the sciences, industry, arts, and humanities—need to know about ML, and what do effective approaches to teaching and learning look like across these diverse domains?". Indeed, little published research examines what effective approaches and tools can support ML education for non-majors.

A recent study by Sulmont, Patitsas, and Cooperstock [26, 27] attempted to investigate instructors' experiences of teaching machine learning to non-majors in order to identify some effective teaching tactics and also some barriers non-majors face when learning about ML. In two articles published from this study [26, 27], the authors assert that teaching ML to non-majors (with little to no CS and statistics background) is possible. However, they argue that more work is still needed in order to develop educational tools (e.g. visualization) that can support effective learning and teaching of ML for non-majors. In this study, the authors interviewed 10 instructors about non-majors' preconceptions about ML and what they find easy or difficult while learning about ML. One preconception held by non-majors was that ML is not accessible without a sufficient CS and statistics background. Instructors noted that non-majors often feel that they are unfit to implement ML. A preconception debunked in the literature [10, 28]. Sulmont, Patitsas, and Cooperstock also identify some pedagogical tactics the instructors found as effective when teaching ML to non-majors. A popular tactic was the use of visualization, including drag-and-drop interfaces with little to no programming required (e.g. Microsoft Azure Machine Learning Studio). Another tactic strongly promoted by instructors was the use of real-world ML applications when introducing ML to non-majors (e.g. predicting tumors using the Breast Cancer Wisconsin Data Set [9]). The authors argue that relating ML to the real world contextualizes ML's importance which is of great importance when working with students across different disciplines.

From the above study [26, 27], we can see that non-majors might view ML as an intimidating subject to learn. This suggests the need for low-floor ML educational tools that can both scaffold non-majors learning about ML and increase their motivation. We believe that Google Teachable Machine could serve as a low floor high ceiling educational tool to scaffold the learning and teaching of ML for non-majors.

Gil [10] introduced ML topics to non-programmers using a workflow paradigm and a visual interface named WINGS. WINGS is an intelligent semantic system that allows users to create, tune, and compare ML processes. Gil found that users of WINGS who did not have programming skills were able to understand the methods within the workflow, including ML concepts. She proposes that learning basic ML concepts without programming is more approachable for non-majors due to the less time and effort investment (i.e. no required lessons on programming).

Fiebrink [8] argues for the importance of teaching ML to creative practitioners, including artists and musicians, and highlights the need for more research about this teaching. In her article, she attempts to lay a foundation for the research and practice of ML education in creative domains and beyond. She proposes a set of learning objectives (LO) for students aiming to create creative artifacts with ML, including:

LO1. Understand the structure of supervised learning problems and the capabilities of supervised learning algorithms

LO4. Apply knowledge of ML workflows and practical skill with an existing ML tool/library to create a ML model

LO8. Understand ways ML has been used in other creative work, and draw on this to contextualise one's own work

Fiebrink also proposes a set of teaching strategies to support the learning and teaching of ML for such students. These strategies include: teach appropriate abstractions, use technologies appropriate to creators, integrate creative perspectives, and support experiential learning of ML. Fiebrink describes how these objectives and strategies can be integrated into the design of technologies that can support ML education for creative practitioners.

Our work builds on Fiebrink's work [8] in which we try to investigate the extent to which a popular novice-friendly ML tool, Teachable Machine, can support the objectives and strategies of ML education, such as those proposed by Fiebrink. However, our work differs in the targeted population as we target non-majors of disciplines beyond the Arts.

Indeed, more research is needed on ML education for non-majors. We hope our work will contribute to informing both the design and use of interactive machine learning tools that could support the teaching and learning of ML for non-majors.

Interactive Machine Learning (IML)

While we aim to study the effectiveness of teaching ML to novices using Interactive Machine Learning (IML) tools, it becomes essential to study the design of these tools to understand the scope and gaps present in effectively transferring knowledge. IML is one approach to create ML models. This approach is usually supported by ML tools that promote the idea of "ML for everyone" [29]. Using IML, a user engages in an iterative process in which he/she provides examples, trains a model, tests the model on new examples, and then refines the model using additional training examples [8]. John J. Dudley et al. [5] describe IML as "an interaction paradigm in which a user or user group iteratively builds and refines a mathematical model to describe a concept through iterative cycles of input and review". They state that the main difference between IML and classical models is that in IML, the user is in charge of the interaction to get the desired output from the model [5]. The IML was designed to enable novices in the field of machine learning to use their domain knowledge to predict patterns in their relevant fields using data-sets. This review paper had studied and identified four key aspects, which though are not mandatory in an IML model, but are widely adopted. These aspects include sample review which is the visualization of output samples at instance level, feedback assignment which is the creation or correction of samples to improve match, model inspection that refers to the evaluation of the model quality and coverage, and task overview which is to view the current status of training data coverage and study improvement potential with respect to cost. The workflow of the IML model includes [5]: (1) feature selection, (2) model selection, (3) model steering, (4) quality assessment, (5) termination assessment, and (6) transfer. The principles for designing an IML interface are found to be oriented towards making the goals and constraints of the model explicit, supporting the user to understand the model, focus on the intent, provide representations, promote interactions and keep the user engaged [5].

The advantage of using an IML model over the Classical Machine Learning (CML) model is that the technical knowledge required to choose the features for an ML model is reduced and the process is made quicker and interactive [7]. According to Jerry Alan Fails and Dan R. Olsen, Jr., in this approach, the designer teaches the classifier and the classifier performs feature selection based on feedback. The Crayons tool developed by the authors helped bridge the knowledge gap in the UI designers in the underlying image processing and machine learning techniques. The IML tool is seen to work efficiently with limited training data and a few chosen features at a much faster rate.

Giselle Nodalo, Jose Ma. Santiago III, Jolene Valenzuela, and Jordan Aiko Deja had proposed to develop a platform for novices in machine learning to explore and perform Machine Learning experiments [21]. The participants in the survey they performed had given a rating for different guidelines to be included in their proposed IML tool (using sandbox). It aligned with Jakob Nielsen's and Rolf Molich's [20, 21] heuristics for evaluating interfaces. The participants had shown a strong desire for inclusion of indication of system status for each interaction, visibility, and feedback in the IML tool.

A web-based tool called CHISSL [3] was designed by John J. Dudley and Per Ola Kristensson to lie in between Active Learning (AL) [23] and Visual Interactive Labeling (VIL) [4]. AL gives the user a well-defined task for eliciting labels for the classifier while VIL leaves it open-ended giving the user a scope to explore and understand the labeling process. However, the interface is simple in the case of AL and complex in the case of VIL. The CHISSL's "tip of the iceberg" design provides the user select instances for a significant portion in the screen without including overwhelming plots. The transduction algorithm based on feedback is representation-free as it does not refer to the underlying feature matrix thereby making the process rapid. Both computational and empirical evaluation results are in favor of the model for being fast and accurate [3].

Amershi et al. claim that effective IML systems should be "rapid, focused, and incremental" [2]. Few threats to consider while designing such models include high latency rate, over-abstraction, overspecialization, clutter, and over-plotting [3].

These experiments and results conducted over time indicate that significant improvements in teaching ML can be achieved by altering and improving the design of the IML tools. In this study, we seek to understand and analyze the design of Google's Teachable Machine and its effectiveness as an IML tool in imparting ML concepts to novices.

METHOD

Study Setting

This qualitative study was conducted as a moderated online study. Participants recruitment, study administration, and data collection occurred over videoconferencing and screen sharing. The authors, as moderators in this online study, administered the study and collected qualitative data by observing participants' behavior while they used Google Teachable Machine and by interviewing those participants.

The participants in this study used Google Teachable Machine on the web using their personal computers. Each participant built two ML models that classify images. The participants used their webcams to collect data for the two designated ML models.

Screen sharing was used to observe how participants used Google Teachable Machine to collect training data using their webcams, to create classes, to label the data, and to evaluate and then refine ML models. Videoconferencing was used to observe participants' behavior while they collect data for their ML models using their webcams. Also, we used videoconferencing to observe participants' body language during the study and to conduct the interviews. Both videoconferencing and screen sharing were recorded.

Participants

Four non-major undergraduate students with no background in ML and no to limited background in CS and statistics were recruited, representing the population of interest described in this study. Two participants are male students and two are female students. The first participant is a 19-year-old male undergraduate student majoring in electrical and electronic engineering. The second participant is a 23-year-old female who recently graduated from medical school. The third participant is a 19-year-old male student majoring in Medicine. The fourth participant is a 23-year-old female who recently graduated with a major in electronics and communication engineering.

The Design of Google Teachable Machine

Google Teachable Machine allows the users to feed data in the form of images by using the webcam of their computers. The input data gathered can be then stored under a label and used for training the model. There can be as many classes included as the user wants. Training is done in the browser and stays on the computer without getting stored or transmitted elsewhere. Finally, the obtained model is tested and tweaked to improve the performance indicated by an increase in the confidence score. The built model can then be stored in the drive for other uses. This adopts a supervised learning approach where the model is trained using a known set of inputs and their corresponding outputs. Google Teachable Machine not only accepts data in the form of images, but it can also be fed with sounds and poses as inputs. However, in this study, the participants will work with image data only.

Study Procedure and Tasks

In order to understand how Google Teachable Machine shapes non-majors' understanding of ML and how they use it to build, evaluate, and refine ML models, we conducted a small-scale qualitative user study with four non-major participants who trained two ML models to classify images using Google Teachable Machine. We held a 45-minute synchronous online session over videoconferencing and screen sharing with each participant. During the session, each participant trained two ML models using Google Teachable Machine and we conducted two interviews with him/her. In this section, we describe the structure of the 45-minute session and the tasks participants engaged in during the study.

Once the session started, we conducted the first interview with the participants. We asked them to introduce themselves, including their names, ages, and majors. After this introduction, we asked questions related to whether participants held any preconceptions about ML.

After this first interview, the participants began with an interactive tutorial on how to train an ML model that classifies different hand gestures using Google Teachable Machine, which lasted 10 minutes. This step-by-step interactive tutorial starts with fixed three classes and does not allow for data labeling. During this tutorial, the participants engaged in collecting data for the three model classes using their webcam. By the end of this tutorial, the participants were able to train an ML model that recognizes three different hand gestures. In the follow-up task, we asked the participants to train the same hand gestures ML model using the full version of Google Teachable Machine in which they can engage in data labeling and can specify the number of classes needed for their ML models.

Then, the participants started training a bigger-scale ML model that we call the face-touching ML model. This ML model should recognize participant's face and whether he/she is touching his/her face or not. Before they start training the face-touching model with Google Teachable Machine, we asked the participants to describe how his/her final face-touching models should work. After developing a theory of how the face-touching model should work, the participants started training the face-touching model using Google Teachable Machine. The participants engaged in collecting training data using their webcams, creating classes, labeling data, evaluating the accuracy of the model, and refining the model to reach a higher confidence score. During this process, the participants were exposed to ML concepts such as overfitting, underfitting, data labeling, and model evaluation. Also, they were exposed to aspects related to epochs, sample size, sample versatility, and negative examples. After completing the face-touching model training task, a second interview was conducted with each participant.

Data Collection Methods

During the 45-minute session with each participant in this qualitative user study, we collected data through observation and interviews. We recorded both the videoconferencing and the screen sharing for the entire session. An observation was conducted for the entire 45-minute session. We also collected self-reported data by conducting two structured interviews with each participant. The first interview took place before the participants started using Google Teachable Machine, and the second interview took place once the participants finished training two ML models using Google Teachable Machine.

Observation

The observation involved observing participant's behavior over videoconferencing and observing how he/she uses and navigates Google Teachable Machine over screen sharing. Over videoconferencing, we observed participant's body language while using Google Teachable Machine and how he/she engages in the process of collecting the training data for their ML models using a webcam. Over screen sharing, we observed how participants used Google teachable Machine while training the two designated ML models.

Structured Interviews

We collected self-reported data by conducting two structured interviews over videoconferencing. The process of developing open-ended interview questions was informed by pedagogical content knowledge (PCK) literature on ML education for non-majors [8, 27]. A total of four non-majors were interviewed before and after they used Google Teachable Machine to train ML models.

The first interview was designed to collect data related to participants' preconceptions about ML. This first interview took place at the beginning of the session and before the participants started using Google Teachable Machine. The questions in this interview were: Are you familiar with the concept of machine learning? If yes, could you explain your understanding of it? What do you think about designing a machine learning model? Do you think it would be easy or challenging? How likely are you to take a formal course in machine learning in the future?

Do you think ML will be useful in your area of study? If yes, could you explain your thoughts on it?

The second interview took place after the participants completed training the two designated ML models with Google Teachable Machine. This second interview included questions related to training the face-touching model. We asked the participants the following questions: Can you explain the process you went through while training the face-touching ML model? How did your initial face-touching model evaluate? What kind of refinement did you make to your initial model? Did it perform better after this refinement? Which parts did you find the most challenging while training the face-touching model?

Additional second interview questions were: Based on your understanding, can you explain the process behind machine learning? What do you think about designing a machine learning model? What do you think about Google Teachable Machine user interface? Can you describe any challenges you encountered while using Google Teachable Machine to train ML models? What would you ideally like your machine learning model to do? What features would you like to add? Do you think ML will be useful in your area of study? If yes, could you explain your thoughts on it? How likely are you to take a formal course in machine learning in the future?

All first and second interviews with each of the four participants were video recorded, and a transcription of all interviews was conducted.

Analysis

Grounded theory approach is used to qualitatively analyze the data gathered from the interviews conducted and observations made through videoconferencing and screen sharing and the video recordings of both the videoconferencing and screen sharing. The sequence of steps adopted to conduct this analysis includes open coding, axial coding, selective coding, and comparative analysis [19]. From this model, we study the effectiveness of Google Teachable Machine in helping the participants understand ML concepts and the ease of use of the interface which might lead to better design specifications in the future with further research. Using Microsoft Excel, the coding tasks were performed and a diagram representing the findings from the analysis was drawn. The steps followed are discussed below.

Open Coding

This step involved the transition of gathered data into codes by the two authors of this paper. Codes represent significant actions, responses, interactions, events, etc. [25]. The open coding resulted in 40 codes. For instance, in the response below, a participant explains her understanding of ML after using Google Teachable Machine. This response is coded as an "improved ML mental model" as it indicates a change in her initial mental model of the concepts of ML.

> Participant 4: "Machine learning is the process of building a model by using sample data in order to train the model to obtain results: predictions/output rather than manually programming the same"

Also, in the example below, the participant encounter difficulties in understanding the learning rate concept. This response is coded as "difficulty in understanding learning rate concept".

> Participant 1: "I don't understand how this learning rate works. What does it do?"

Reliability Analysis

Inter-code reliability analysis [15] was performed, and Cohen's Kappa [17] was calculated to measure the coherence in the coding. It was found to be 0.87 which indicated a strong to an almost perfect level of agreement. Discussions were conducted to resolve any case of discordance.

Axial Coding

This step involved refining the previous step by categorizing the codes generated. This helps in observing the phenomena referred to as "repeated patterns of events, happenings, actions, and interactions that represent people's responses to the problems and situations, which they encounter in the social context." [25]

Fourteen categories were developed as a result of axial coding. "Productive/positive tinkering", for example, refers to behaviors related to exploration and learning through trial and error such as "changing epochs value" and "adding more image samples".

Selective Coding

This step of coding involved the integration of codes and their categories by identifying their relationship in terms of a story. This gives an overview of the essence of the analysis. We used diagramming in this step and it has resulted in creating six different stories.

Comparative Analysis

One phenomenon that emerged from our data was that participants' interest in learning about ML and their perceptions about the approachability of ML was affected by exposure to ML. We conducted comparative analysis to look at how participants' interest in and perceptions about ML before any exposure to ML differed from their interest in and perceptions about ML after an exposure to ML using Google Teachable Machine. For example, a comparative analysis of participants' data before and after using Google Teachable Machine revealed that before exposure to ML, the participants showed no interest in learning about ML while after an exposure to ML using Google Teachable Machine, the participants showed more interest in learning about ML.

RESULTS

In our analysis, we identified different themes that provide answers to how Google Teachable Machine shapes non-majors' understanding of basic ML concepts and how they use Google Teachable Machine, as an IML tool, to train ML models. We identified themes related to: (1) how Google Teachable Machine shaped participants' ML mental models, (2) how Google Teachable Machine helped participants train, evaluate, and refine ML models, and (3) how using Google Teachable Machine to learn about ML affected participants' perceptions, interest, and attitudes towards learning about ML.

Mental Models of ML Concepts

The participants were able to construct a viable understanding of ML after using Google Teachable Machine. All participants were able to describe the concept of machine learning. They described ML as follows:

> Participant1: “The process of training a model to take decisions from previous experiences”
>
> Participant2:"Training a computer to recognize new events or inputs based off data it has already received of those inputs"

The two interviews we had with each participant, one before they use Google Teachable Machine and another after they used Google Teachable Machine, revealed the effect of using Google Teachable Machine on constructing participants' mental model of the concept of ML. For instance, when we asked participant 4 before she starts using Google Teachable Machine if she is familiar with the concept of ML, her answer was:

> Participant 4:"No, except for the basic idea that it is an AI application which enables processes to learn and improve by experience and training over specific configuration/programming"

We asked her again after she used Google Teachable Machine to train two ML models if she could explain her understanding of ML. Her answer was:

> Participant 4:"Machine learning is the process of building a model by using sample data in order to train the model to obtain results: predictions/output rather than manually programming the same"

The difference between her two answers shows how using Google Teachable Machine has helped her construct a more viable understanding of how ML works as she was able to explain what ML is in more details.

In addition to constructing a mental model of the concept of ML, the participants were able to construct mental models of the epoch concept and the learning rate concept. Initially, the participants found these concepts challenging. However, tinkering with the options related to these concepts, which is facilitated by the intuitive Google Teachable Machine user interface, and the instant feedback that Google Teachable Machine provides about the model performance facilitated learning about these concepts through trial and error. Ultimately, the participants were able to explain the epoch and learning rate concepts and how they affect their ML models. Below is an example of how a participant described her understanding of epoch and learning rate concepts after tinkering (changing the values) and trial and error with Google Teachable Machine:

> Participant 4: "The number of iterations it takes . . . more iterations, the better it learns “, (explaining epochs)
>
> Participant 4: "The smaller the learning rate, the better the model gets trained. As I increase the learning rate, it was rather slow and it reached a point where it failed to recognize between the first and the second command".

ML models Training

Google Teachable Machine facilitated training accurate ML models by the non-major participants. All participants were able to train models with 90-100 percent of accuracy. Mainly as a result of the simple and intuitive Google Teachable Machine user interface and the instant feedback that Google Teachable Machine provides that helped the participants evaluate the accuracy of their models on a 100 percent scale.

The participants refined and increased the accuracy of their models by: (1) increasing the number of image examples they add to a model class, (2) improving how they capture the image examples and collect diverse examples, or (3) changing the epoch, learning rate, and batch values. For example, one participant refined his model by changing epochs value from 50 to 100 which resulted in a more accurate model. Below is an example of how a participant went through the process of assessing and refining his model. In this example, participant 3 was trying to train a hand-gesture model with one class for waving hand and another class for sitting still with no hand gestures:

> (Participant 3 testing his hand-gestures model. Model showing inaccurate results. Participant 3 recognized the inaccuracy of the model)
> Moderator:” is it working?”
> Participant 3:”ah, kind of .. except class 2 is like doesn't say a 100 percent like it does for class 1 .. it says 50 50 . . . maybe if I put in more images?”
> (He opened his webcam to take image examples, then adjusted his posture then started taking images of himself waving his hand, then started moving his hand around while waving it and taking pictures)
> (He retrained the model. The model showing accurate results)
> Moderator: ”so, do you think it's accurate now?”
> Participant 3:”yeah!”

In the above example, using Google Teachable Machine, participant 3 was able to evaluate the accuracy of his model then refine it by adding more image examples and improving the way he captures image examples and provides diverse examples for one class. Then, he was able to see how the refinements affected his model accuracy.

Perceptions, Interest, and Attitudes Towards ML

Our interviews with the participants revealed that the non-major participants perceive ML as important in their fields of study. They were able to connect ML to their fields and give examples of how they want to use ML in their fields. For instance, we asked the participants if they thought ML will be useful in their areas of study, their responses were as follow:

> Participant 1: "yes, in the area of load forecasting and smart grids"
> Participant 2: "Yes". Then, she proceeds to give an example: "when I speak something or give a command, the model is supposed recognize it and show pictures or any slides for teaching purposes"
> Participant 3: "Yes, as an aspiring doctor I can see the benefits of being able to train a computer to recognize

> patterns in patients. It could lead to more efficient and accurate diagnoses in the future"
>
> Participant 4:"Although it might not be directly relevant, it sure plays a huge role in the working of the products in the SaaS industry I'm in right now and would help with further enhancements/suggestions"

In addition, our data analysis revealed that using Google Teachable Machine to introduce ML to the non-major participants increased their interest in ML, helped them perceive ML as more approachable, and resulted in positive attitudes towards learning about ML. In our initial interviews with the participants, they stated that they are not interested in learning about ML and that they perceive ML as challenging. However, in our second interviews with them after they used Google Teachable Machine to train ML models, they stated that they became more interested in learning about ML, and that they now perceive ML as more approachable. Below is an example of one participant's answers about their interest in learning about ML before and after they used Google Teachable Machine to train ML models:

> First interview: Moderator: "How likely are you to take a formal course in machine learning in the future?"
>
> Participant: "Not very likely"
>
> Second interview: Moderator: "How likely are you to take a formal course in machine learning in the future?"
>
> Participant: "This got me interested, I would most definitely want to experiment further. With either a formal professional course or one online, to say the least"

One example of how participants' perceptions about the approachability of ML have changed after learning about ML with Google Teachable Machine is:

> First interview: Moderator: "How approachable do you think is designing a machine learning model?"
>
> Participant 4:"Moderately challenging"
>
> Second interview: Moderator: "Which parts did you find the most challenging while training the face-touching model?"
>
> Participant 4:"It was fairly simple and very interesting. No specific challenges"
>
> Moderator: "How was your experience designing a machine learning model on your own?"
>
> Participant 4:"It was great fun, very interesting too"

Finally, after using Google Teachable Machine to train ML models, the participants held positive attitudes towards learning about ML. The participants used positive terms to describe their experience learning about ML using Google Teachable Machine. For instance:

> Participant 1: "It was interesting"
>
> Participant 2: "It was good"
>
> Participant 3: "that's pretty cool", "It was fun to see how machine learning worked"
>
> Participant 4:"It was great fun, very interesting too"

DISCUSSION

In this study, we attempted to study non-majors' learning about ML with Google Teachable Machine, an IML tool. The scarce literature on ML education for non-majors shows that non-majors are able to understand and implement ML [26, 27, 10, 28]. Our results add to this body of literature as it also shows non-majors' ability to understand and apply basic ML concepts. Using Google Teachable Machine, the participants were able to apply supervised learning in training ML models with 90-100 percent of accuracy. They were able to employ advanced concepts such as learning rate, epochs, and batch size to increase the accuracy of their ML models.

In our study, we proposed introducing ML to non-majors using Google Teachable Machine because of its IML design that engages users in an iterative process of providing examples, training a model, testing the model on new examples, and then refining the model [8]. Our findings illustrate how this design significantly affected how participants were able to construct viable mental models of ML concepts and how they were able to train accurate ML models. The design of Google Teachable Machine allowed the participants to engage in an iterative process of collecting examples, training, then receiving and instant feedback from Google Teachable Machine about the accuracy of the model. This helped the participant to engage in productive tinkering and learning through trial and error and has eventually resulted in improving participants' mental models of basic ML concepts.

Our initial interviews with the participants which took place before introducing them to ML revealed how non-majors perceive ML as important within their fields, and how they were able to describe viable applications of ML within their fields. However, they did not show an interest in learning about ML and they perceived ML as challenging, which aligns with the results of [26, 27] that show how non-majors hold a preconception about ML as not accessible without a sufficient CS and statistics background. Interestingly, our post interviews which took place after learning about ML with Google Teachable Machine reveal an increased interest in learning about ML and more notions of ML as interesting and approachable. In discussing these results, we believe that the initial disinterest in ML might be a result of holding a preconception about ML as a challenging subject, and that using an intuitive and user-friendly IML tool, Google Teachable Machine, has facilitated non-majors' learning about ML, thus resulting in perceiving ML as a less intimidating subject.

Although Google Teachable Machine provides non-majors with introductory learning about ML and might not cover advanced ML concepts, we believe that it represents an efficient tool to introduce non-majors to ML because of its positive effect on non-majors' interest in ML, on their attitudes towards learning about ML, and on the construction of viable mental models of basic ML concepts. All these effects are crucial when it comes to novices' learning as it would crucially affect how they proceed with their learning and how viable this learning would be.

Implications For Research and Design

As this study shows how non-majors are able to understand and apply ML concepts and how they perceive the importance of ML applications in their fields, more research needs to be done on the learning and teaching of ML for non-majors. Most importantly, more research needs to focus on how to provide ML education that is accessible to non-majors which would help scaffold their learning and increase their motivation to learn more advanced applications of ML. This would lead to the creation of innovative ML applications in their fields of study.

The positive results of using Google Teachable Machine in our study encourages more research on the design and study of similar ML tools in the area of ML education for non-majors. As Google Teachable Machine covers basic ML concepts, further research needs to study other ML tools that cover more advanced topics. Also, further research could investigate non-majors' transitioning from learning with simple IML tools to learning with more advanced ML tools. This would allow studying whether IML tools would positively or negatively impact non-majors learning about ML as the proceed to more advanced ML education.

Limitations

A key limitation in this study is the small sample size. Due to COVID-19 disruption, only four participants were recruited in this study. Another limitation is participants' majors. Only two majors were involved in this study, Medicine and Electrical and Electronic Engineering, limiting the generalizability to the larger non-majors population.

CONCLUSION

As ML is a relatively new field, research on the teaching and learning of ML is scarce, especially for non-majors. In this qualitative study on machine learning education for non-majors, we attempted to investigate the effectiveness of Google Teachable Machine, an interactive machine learning tool, in helping non-majors understand and effectively apply basic ML concepts.

Our findings reveal that Google Teachable Machine facilitates the construction of viable mental models of basic ML concepts and that non-majors can train highly accurate ML models using Google Teachable Machine. Non-majors were able to construct mental models of the concept of ML and concepts like epoch and learning rate. Initially, the participants found these concepts challenging but the features supported by the design of Google Teachable Machine facilitated learning about these concepts. In addition, Google Teachable Machine positively affected non-majors' perceptions, interest, and attitudes towards learning about ML. Our data analysis revealed that using Google Teachable Machine to introduce ML to non-majors increased participants' interest in ML, helped them perceive ML as more approachable, and resulted in positive attitudes towards learning about ML.

We hope that our work contributes to encouraging and informing the design and use of IML tools that could support the teaching and learning of ML for non-majors.

REFERENCES

[1] Runway AI. 2020. RunwayML. (2020). Retrieved February 16, 2020 from `https://runwayml.com/`

[2] Saleema Amershi, Maya Cakmak, William Bradley Knox, and Todd Kulesza. 2014. Power to the people: The role of humans in interactive machine learning. *Ai Magazine* 35, 4 (2014), 105–120.

[3] Dustin Arendt, Emily Saldanha, Ryan Wesslen, Svitlana Volkova, and Wenwen Dou. 2019. Towards rapid interactive machine learning: evaluating tradeoffs of classification without representation. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. 591–602.

[4] Jürgen Bernard, Marco Hutter, Matthias Zeppelzauer, Dieter Fellner, and Michael Sedlmair. 2017. Comparing visual-interactive labeling with active learning: An experimental study. *IEEE transactions on visualization and computer graphics* 24, 1 (2017), 298–308.

[5] John J Dudley and Per Ola Kristensson. 2018. A review of user interface design for interactive machine learning. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 8, 2 (2018), 1–37.

[6] Steven D Essinger and Gail L Rosen. 2011. An introduction to machine learning for students in secondary education. In *2011 Digital Signal Processing and Signal Processing Education Meeting (DSP/SPE)*. IEEE, 243–248.

[7] Jerry Fails and Dan Olsen. 2003. Interactive Machine Learning. *International Conference on Intelligent User Interfaces, Proceedings IUI* (05 2003). DOI: `http://dx.doi.org/10.1145/604045.604056`

[8] Rebecca Fiebrink. 2019. Machine Learning Education for Artists, Musicians, and Other Creative Practitioners. *ACM Trans. Comput. Educ.* 19, 4, Article Article 31 (Sept. 2019), 32 pages. DOI: `http://dx.doi.org/10.1145/3294008`

[9] Center for Machine Learning and Intelligent Systems. 2020. Breast Cancer Wisconsin (Diagnostic) Data Set. (2020). Retrieved February 16, 2020 from `https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)`

[10] Yolanda Gil. 2016. Teaching big data analytics skills with intelligent workflow systems. In *Thirtieth AAAI Conference on Artificial Intelligence*.

[11] Google. 2019. Google Teachable Machine. (2019). Retrieved February 16, 2020 from `https://teachablemachine.withgoogle.com/`

[12] Tom Hitron, Yoav Orlev, Iddo Wald, Ariel Shamir, Hadas Erel, and Oren Zuckerman. 2019. Can Children Understand Machine Learning Concepts? The Effect of Uncovering Black Boxes. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, Article Paper 415, 11 pages. DOI: `http://dx.doi.org/10.1145/3290605.3300645`

[13] Tom Hitron, Iddo Wald, Hadas Erel, and Oren Zuckerman. 2018. Introducing Children to Machine Learning Concepts through Hands-on Experience. In *Proceedings of the 17th ACM Conference on Interaction Design and Children (IDC '18)*. Association for Computing Machinery, New York, NY, USA, 563–568. DOI:http://dx.doi.org/10.1145/3202185.3210776

[14] Amy J. Ko. 2017. We need to learn how to teach machine learning. (Feb 2017). Retrieved February 16, 2020 from https://medium.com/bits-and-behavior/we-need-to-learn-how-to-teach-machine-learning-acc78bac3ff8

[15] Karen S Kurasaki. 2000. Intercoder reliability for validating conclusions drawn from open-ended interview data. *Field methods* 12, 3 (2000), 179–194.

[16] Radu Mariescu-Istodor and Ilkka Jormanainen. 2019. Machine Learning for High School Students. In *Proceedings of the 19th Koli Calling International Conference on Computing Education Research (Koli Calling '19)*. Association for Computing Machinery, New York, NY, USA, Article Article 10, 9 pages. DOI:http://dx.doi.org/10.1145/3364510.3364520

[17] Mary L McHugh. 2012. Interrater reliability: the kappa statistic. *Biochemia medica: Biochemia medica* 22, 3 (2012), 276–282.

[18] Microsoft. 2020. Microsoft Azure Machine Learning Studio. (2020). Retrieved February 16, 2020 from https://azure.microsoft.com/en-us/services/machine-learning/

[19] Bilge Mutlu and Jodi Forlizzi. 2008. Robots in organizations: The role of workflow, social, and environmental factors in human-robot interaction. *2008 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (2008), 287–294.

[20] Jakob Nielsen and Rolf Molich. 1990. Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 249–256.

[21] Giselle Nodalo, Jose Ma. Santiago, Jolene Valenzuela, and Jordan Aiko Deja. 2019. On Building Design Guidelines for An Interactive Machine Learning Sandbox Application. In *Proceedings of the 5th International ACM In-Cooperation HCI and UX Conference (CHIuXiD'19)*. Association for Computing Machinery, New York, NY, USA, 70–77. DOI:http://dx.doi.org/10.1145/3328243.3328253

[22] Alexander Scheidt and Tim Pulver. 2019. Any-Cubes: A Children's Toy for Learning AI: Enhanced Play with Deep Learning and MQTT. In *Proceedings of Mensch Und Computer 2019 (MuC'19)*. Association for Computing Machinery, New York, NY, USA, 893–895. DOI:http://dx.doi.org/10.1145/3340764.3345375

[23] Burr Settles. 2009. *Active learning literature survey*. Technical Report. University of Wisconsin-Madison Department of Computer Sciences.

[24] R. Benjamin Shapiro and Rebecca Fiebrink. 2019. Introduction to the Special Section: Launching an Agenda for Research on Learning Machine Learning. *ACM Trans. Comput. Educ.* 19, 4, Article Article 30 (Oct. 2019), 6 pages. DOI:http://dx.doi.org/10.1145/3354136

[25] Anselm Strauss and Juliet Corbin. 1990. *Basics of qualitative research*. Sage publications.

[26] Elisabeth Sulmont, Elizabeth Patitsas, and Jeremy R. Cooperstock. 2019a. Can You Teach Me To Machine Learn?. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education (SIGCSE '19)*. Association for Computing Machinery, New York, NY, USA, 948–954. DOI:http://dx.doi.org/10.1145/3287324.3287392

[27] Elisabeth Sulmont, Elizabeth Patitsas, and Jeremy R. Cooperstock. 2019b. What Is Hard about Teaching Machine Learning to Non-Majors? Insights from Classifying Instructors' Learning Goals. *ACM Trans. Comput. Educ.* 19, 4, Article Article 33 (July 2019), 16 pages. DOI:http://dx.doi.org/10.1145/3336124

[28] Thomas Way, Lillian Cassel, Paula Matuszek, Mary-Angela Papalaskari, Divya Bonagiri, and Aravinda Gaddam. 2016. Broader and Earlier Access to Machine Learning. In *Proceedings of the 2016 ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE '16)*. Association for Computing Machinery, New York, NY, USA, 362. DOI:http://dx.doi.org/10.1145/2899415.2925485

[29] Qian Yang, Jina Suh, Nan-Chen Chen, and Gonzalo Ramos. 2018. Grounding interactive machine learning tool design in how non-experts actually build models. In *Proceedings of the 2018 Designing Interactive Systems Conference*. 573–584.

[30] Kevin Zhu. 2019. *An educational approach to machine learning with mobile applications*. Ph.D. Dissertation. Massachusetts Institute of Technology.

[31] Abigail Zimmermann-Niefield. 2019. Machine Learning Education for Young People without Programming Experience. In *Proceedings of the 2019 ACM Conference on International Computing Education Research (ICER '19)*. Association for Computing Machinery, New York, NY, USA, 359–360. DOI:http://dx.doi.org/10.1145/3291279.3339446

[32] Abigail Zimmermann-Niefield, Makenna Turner, Bridget Murphy, Shaun K. Kane, and R. Benjamin Shapiro. 2019. Youth Learning Machine Learning through Building Models of Athletic Moves. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children (IDC '19)*. Association for Computing Machinery, New York, NY, USA, 121–132. DOI:http://dx.doi.org/10.1145/3311927.3323139